XtreemOS: a Sound Foundation for Cloud Infrastructure and Federations

Thilo Kielmann, Guillaume Pierre, Christine Morin

XtreemOS: a Sound Foundation for Cloud Infrastructure and Federations

Thilo Kielmann, Guillaume Pierre, Christine Morin

Abstract XtreemOS is a Linux-based operating system with native support for virtual organizations (VO's), for building large-scale resource federations. XtreemOS has been designed as a grid operating system, supporting the model of resource sharing among independent administrative domains. We argue, however, that the VO concept can be used to establish either resource sharing or resource isolation, or even both at the same time. We outline XtreemOS' fundamental properties and how its native VO support can be used to implement cloud infrastructure and cloud federations.

1 XtreemOS

Developing and deploying applications for traditional (single computer) operating systems is well understood. Federated resources like in grid environments, however, are generally perceived as highly complex and difficult to use. The difference lies in the underlying system achitecture. Operating systems provide a well-integrated set of services like processes, files, memory, sockets, user accounts and access rights. Grids, in contrast, add a more or less heterogeneous middleware layer on top of the operating systems of the federated resources. This lack of integration has lead to a lot of complexity, for both users and administrators.

To remedy this situation, XtreemOS [7] has been designed as a *grid operating system*. While being based on Linux, it provides a comprehensive set of services as well as a stable interface for wide-area, dynamic, distributed infrastructures composed of heterogeneous resources spanning multiple administrative domains. The fundamental issues addressed by XtreemOS are scalability and transparency.

Thilo Kielmann and Guillaume Pierre
Vrije Universiteit, Amsterdam, The Netherlands, e-mail: kielmann@cs.vu.nl, gpierre@cs.vu.nl

Christine Morin
INRIA, Centre Rennes - Bretagne Atlantique, Rennes, France, e-mail: Christine.Morin@irisa.fr

Scalability. Wide-area, distributed infrastructures like grids easily consist of thousands of nodes and users. Along with this scale comes heterogeneity of (compute and file) resources, networks, administrative policies, as well as churn of resources and users. XtreemOS addresses these issues by its integrated view on resources, along with its built-in support for virtual organizations (VO's) that provide the scoping for resource provisioning and access. For sustained operation, XtreemOS provides an infrastructure for highly-available services, to support both its own critical services and user-defined application services.

Transparency. Vital for managing the complexity of grid-like infrastructures is providing transparency for the distributed nature of the environment, by maintaining common look–and–feel for the user and by exposing distribution and federation only as much as necessary. To the user, XtreemOS provides single sign-on access, Linux look–and–feel via grid-aware shell tools, and API's that are based on both POSIX and the Simple API for Grid Applications (SAGA). For the administrators of VO's and site resources, XtreemOS provides easy-to-use services for all management tasks.

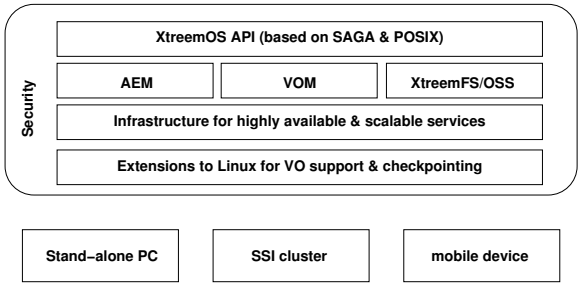

Fig. 1 The XtreemOS system architecture

Figure 1 summarizes the XtreemOS system architecture. XtreemOS comes in three flavours; one for stand-alone nodes (PC's), one for clusters providing a single-system image (SSI), and one for mobile devices. Common to all three flavours are the Linux extensions for VO support, providing VO-based user accounts via kernel modules [1]. PC and cluster flavour also share support for grid-wide, kernel-level job checkpointing.

The infrastructure for highly available and scalable services consists of implementations of distributed servers and of virtual nodes [6]. The *distributed servers* form a transparent group of machines that provide their services through a shared (mobile IPv6) address. Within the group, load balancing and fault tolerance are implemented transparent to the clients. The *virtual nodes* provide fault-tolerant service replication via a Java container, transparent to the service implementation itself.

Central to VO-wide operation are the services AEM, VOM, the XtreemFS file system and the OSS mechanism for sharing volatile application objects. The VO management services (VOM) provide authentication, authorization, and accounting

for VO users and resources. VO's can be managed dynamically through their whole life cycle while user access is organized with flexible policies, providing customizable isolation, access control, and auditing. The VO management services, together with the kernel modules enforcing local accounts and policies provide a security infrastructure underlying all XtreemOS functionality.

The Application Execution Management (AEM) relies on the Scalaris [4] peer–to–peer overlay among the compute nodes of a VO that allows to discover, select, and allocate resources to applications. It provides POSIX-style job control to launch, monitor, and control applications.

The XtreemFS grid file system [2] provides users with a global, location independent view of their data. XtreemFS provides a standard POSIX interface, accomodating from multiple VO's, across different administrative domains. It provides autonomous data management with self-organized replication and distribution. The Object Sharing Service (OSS) provides access to volatile, shared objects in main memory segments.

The XtreemOS API's accomodate existing Linux and grid applications, while adding support to XtreemOS' unique features. POSIX interfaces support Linux applications; grid-aware shell tools seemlessly integrate compute nodes within a VO. Grid applications find their support via the OGF-standardized Simple API for Grid Applications (SAGA) [5]. API's for XtreemOS-specific functionality (XtreemOS credentials, AEM's resource reservation, XtreemFS URL's, OSS shared segments, etc.) are provided as SAGA extension packages, commonly referred to as the XOSAGA API.

2 Cloud Infrastructure and Federations

Grid infrastructures operate by sharing physical resources among the users of a VO; sharing and isolation are managed by the site-local operating systems and the VO-wide (middleware) services. Although cloud computing as such is still in its infancy, the *Infrastructure as a Service* paradigm (IaaS) has gained importance. Here, virtualized resources are rented to cloud users; sharing and isolation are managed by the Virtual Machine Managers (VMM's). What makes this model attractive is that users get full control over the virtual machines, while the underlying IaaS infrastructure remains in charge of resource sharing and management. An important drawback of this model is that it provides only isolated machines rather than integrated clusters with secure and fast local networks, integrated user management and file systems.

This is where XtreemOS provides added value to IaaS clouds [3]. Figure 2 shows how XtreemOS can integrate resources from one or more IaaS providers to form a clustered resource collection for a given user. Within a single IaaS platform, XtreemOS integrates multiple virtual machines similar to its SSI cluster version, to form a cloud cluster with integrated access control based on its VO-management mechanisms, here applied to a user-defined, dynamic VO. Across multiple IaaS platforms, the same VO management mechanisms allow the federation of multiple cloud

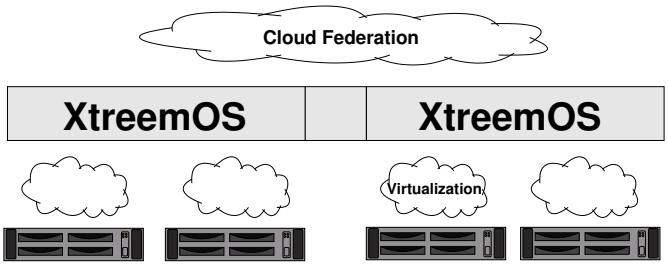

Fig. 2 XtreemOS integrating IaaS resources

clusters to a user's VO. In combination with the XtreemFS file system, such IaaS federations provide flexibly allocated resources that match a user's requirements, while giving full control over the virtualized resources.

XtreemOS extends Linux by its integrated support for VO's. Within grid computing environments, VO's enable sharing of physical resources. Within IaaS clouds, VO's enable proper isolation between clustered resources, thus allowing to form unified environments tailored to their users.

Acknowledgements

This work has been supported by the EU IST program as part of the XtreemOS project (contract FP6-033576).

References

1. M. Coppola, Y. Jégou, B. Matthews, Ch. Morin, L.P. Prieto, O.D. Sánchez, E.Y. Yang, H. Yu: Virtual Organization Support within a Grid-Wide Operating System. IEEE Internet Computing, Vol. 12, No. 2, 2008
2. F. Hupfeld, T. Cortes, B. Kolbeck, J. Stender, E. Focht, M. Hess, J. Malo, J. Marti, E. Cesario: The XtreemFS Architecture—a Case for Object-based File Systems in Grids. Concurrency and computation: Practice and experience, Vol. 20, No. 17, 2008.
3. Ch. Morin, Y. Jégou, J. Gallard, P. Riteau: Clouds: a new Playground for the XtreemOS Grid Operating System. Parallel Processing Letters, Vol. 19, No. 3, 2009.
4. T. Schütt, F. Schintke, A. Reinefeld: Scalaris: Reliable Transactional P2P Key/Value Store – Web 2.0 Hosting with Erlang and Java. 7th ACM SIGPLAN Erlang Workshop, Victoria, September 2008.
5. Ch. Smith, T. Kielmann, S. Newhouse, M. Humphrey: The HPC Basic Profile and SAGA: Standardizing Compute Grid Access in the Open Grid Forum. Concurrency and Computation: Practice and Experience, Vol. 21, No. 8, 2009.
6. M. Szymaniak, G. Pierre, M. Simons-Nikolova, M. van Steen: Enabling Service Adaptability with Versatile Anycast. Concurrency and Computation: Practice and Experience, Vol. 19, No. 13, 2007.
7. XtreemOS: www.xtreemos.eu